Measuring Productivity Dispersion

Bartelsman, Eric J.; Wolf, Zoltan

published in
Oxford Handbook of Productivity Analysis
2018

DOI (link to publisher)
10.1093/oxfordhb/9780190226718.013.18

document version
Publisher's PDF, also known as Version of record

Measuring Productivity Dispersion

Eric J. Bartelsman and Zoltan Wolf

The Oxford Handbook of Productivity Analysis
Edited by Emili Grifell-Tatjé, C.A. Knox Lovell, and Robin C. Sickles

Print Publication Date: Sep 2018
Subject: Economics and Finance, Business Economics, Labor and Demographic Economics
Online Publication Date: Sep 2018 DOI: 10.1093/oxfordhb/9780190226718.013.18

Abstract and Keywords

Measuring the dispersion of productivity or efficiency across firms in a market or industry is rife with methodological issues. Nevertheless, the existence of considerable dispersion now is well documented and widely accepted. Less well understood are the economic features and mechanisms underlying the magnitude of dispersion and how dispersion varies over time or across markets. On the one hand, selection mechanisms in both output and input markets should favor the most productive units through resource reallocation, thereby reducing dispersion. On the other hand, innovation and technological uncertainty tend to increase dispersion. This chapter presents a guide to the measurement of dispersion and provides empirical evidence from a selection of countries and industries using a variety of methodologies.

Keywords: productivity, market, industry, technology, innovation, dispersion

18.1. Introduction

HETEROGENEITY in productivity or the efficiency of producers has long been recognized in the academic literature, but traditionally was considered more of a hindrance that needed to be massaged away in analysis, rather than an important feature of economic life requiring theoretical and empirical analysis.[1] Marshall (1920) introduced the notion of a "representative firm" in his "Principles" in order to analyze equilibrium in production. Robbins (1928) notes that Marshall mainly introduced the concept in order to simplify analysis. Robbins then goes on to argue that the construct of the representative firm is not needed for analysis of economic equilibrium and actually may be misleading (Robbins 1928; p.399): "The whole conception, it may be suggested, is open to the general criticism that it cloaks the essential heterogeneity of productive factors—in particular the heterogeneity of managerial ability—just at that point at which it is most desirable to exhibit it most vividly."

Notwithstanding the contribution of Robbins, much of the theoretical work in general equilibrium theory and also in macro theory of business cycles and growth continued to use the representative firm until recently. By contrast, ensuing empirical research eschewed the representative firm (e.g., Farrell 1957; Salter 1960), but did not have theoretical explanations for how productivity differences could coexist. Leibenstein (1966) contrasted deviations from efficiency as described by micro theory (allocative inefficiency), with differences in efficiency across otherwise similar production units. By giving a name to the gap from the most productive firm, "X-inefficiency," Liebenstein may have provided an appealing narrative, but did not satisfy the theoretician's desire for placing the phenomenon in the framework of cost minimization (e.g., Stigler 1976). However, following Stigler's critique and reply (Leibenstein 1978), the path had opened up for future researchers to work on building a framework to understand why (p. 594) productivity dispersion across firms exists and even may be compatible with optimizing behavior in output and input markets.

The explanations generally require some curvature in the profit function of a producer that prevents the most productive firm from selling to all customers in the market. Mechanisms include frictions in the adjustment of factors and the entry and exit of plants, and distortions that drive wedges in the forces pushing toward the equalization of marginal products across plants. Early models of heterogeneous producers that support productivity dispersion in equilibrium are given by Lucas (1978) and Hopenhayn (1992). Other relevant theoretical contributions point the way toward understanding how dispersion may shed light on the measurement of output and inputs (e.g., De Loecker 2011), on frictions in optimization (e.g., Cooper and Haltiwanger 2006), and on distortions to the functioning of markets (e.g., Brown et al. 2016; Hsieh and Klenow 2009). This chapter will provide some guidelines on how to measure productivity dispersion and place it into context of the models.

The recipes given in this chapter for measuring and analyzing dispersion of productivity use longitudinal firm- or plant-level data as collected by statistical agencies in annual production surveys. These data underlie much of the empirical literature reviewed by Bartelsman and Doms (2000) and Syverson (2011). Further, similar data sets are now being explored in empirical studies of productivity, innovation, employment, and trade, for example by the Eurostat ESSNet projects (Bartelsman et al. 2018a), the Organisation for Economic Co-operation and Development (OECD) DynEmp project (Criscuolo et al. 2014), and the European Central Bank CompNet project (Lopez-Garcia and di Mauro 2015). Using these data, the research finds that productivity differences across establishments indeed are large and persistent in all countries, industries, and time periods reviewed.

Dispersion is important as a measure of heterogeneity and also because it is relevant for business dynamism and growth. The role of dispersion for business dynamism and growth has been explored extensively in the context of the relationship between productivity, growth, and reallocation dynamics. A number of papers found that more productive plants are more likely to grow and less likely to exit (recent examples include Foster et al. 2017; Foster et al. 2016a). Another area of application is the frontier literature, which postulates that the technology and practices of the most productive plants, or frontier plants, are adapted by other establishments (e.g., Acemoglu et al. 2006; Bartelsman et al. 2015). In this view, growth is sourced either from innovative activity at the frontier or from the adjustment of nonfrontier establishments, in which they adopt frontier behavior. Yet another area of inquiry is related to the interpretation of dispersion in revenue productivity. Based on the insights in Hsieh and Klenow (2009)—that under certain assumptions about technology and demand, dispersion in productivity reflects market distortions—dispersion in a particular revenue productivity measure has been used to create indicators of misallocation (a recent example is Foster et al. 2016b).

The remainder of the chapter is organized as follows. Before defining productivity and its measures, we start with a theoretical discussion on productivity dispersion. The next section will discuss measurement of productivity at the plant level and will place (p. 595) the simple measures of productivity used in the literature on dispersion in the context of more sophisticated measures discussed in this *Handbook*. Next, some recipes will be provided for computing dispersion measures, taking into account sensitivity to measurement errors. The chapter then will conclude with a review of some evidence on productivity dispersion in a wide variety of industries and countries, as well as thoughts about a model that endogenizes productivity dispersion.

18.2. What Is Productivity Dispersion?

Assume we have an indicator of productivity, ω_{it}, of a production unit i in time period t, that measures how much more, or less, output (in log-points) is produced per unit of input than at some "reference" production unit. This measure of productivity, for a single firm, plant, or decision-making unit, is the basic building block for cross-sectional measures of dispersion (at time t). Dispersion is related to the "width" of the productivity distribution and thus has the same dimensionality as the underlying measure. The empirical distribution of productivity built up from the ω_{it}s that are derived from observed data is the result of our statistical methodology in collecting the data and the computational methods of computing productivity, as well as the result of economic processes driven by decisions made at production units and the interactions between economic agents in input and output markets. Finally, dispersion in productivity can reflect idiosyncracies in the processes driving creation of knowledge and production technology.

In this section we will provide some theoretical background into the drivers of the empirical measure of productivity dispersion. We start with a discussion of statistical issues. Next, we look at two sides of the economic process driving dispersion. First, we look at factors that drive dispersion across firms in their ability to produce output given inputs (i.e., at a certain level of productivity). Second, we look at processes in input and output markets that reallocate inputs and select production units and thus jointly shape the observed productivity distribution. Because of its importance as the building block for measuring dispersion, a separate section is devoted to computation of the relative productivity of a production unit, ω_{it}.

18.2.1. Statistical Issues

Dispersion in productivity is some measure of the distribution of productivity, for example the second moment. The use of the terms *measure, distribution*, and *moment* bring on thoughts about probability and statistics, and possibly about sampling and measurement error. In this section we disentangle statistical issues from the economic phenomenon that we are trying to measure.

(p. 596) From probability theory, we can understand a probability space to consist of a sample space, a set of events, and a function mapping events to a probability. Interpretation of the (empirical) productivity distribution depends on what we think the underlying process is through which outcomes are drawn from the sample space, and how we think about the relationship between events and the available data. For example, we could think of the outcome of NT observations from a longitudinal panel of N firms and T years as being independent draws from a particular sample space and probability mapping.[2] Given the sample size, we could then place error bounds on estimates of the standard error of the probability distribution. Under these assumptions, the interpretation of dispersion of productivity is clear. However, the underlying assumptions may not hold, and deviations require differing interpretations.

To start, the observations may not be independently drawn from the same distribution. This can easily be tested, for example by testing for the equality of the "within" (over time-series dimension) estimate of the standard error with the "between" (over cross-sectional dimension) estimate (see later discussion for details). To our knowledge, the empirical evidence shows that the standard error of the productivity measures across firms in an industry is much larger than the standard deviation of productivity at the firm-level (on average across firms) over time. To distinguish between the two dimensions, we will call the second moment over the cross-section *dispersion* and call the second moment over the time-series of productivity (growth) *volatility*.[3]

Volatility of productivity likely has different "causes" than dispersion of productivity and also plays a distinct role in different types of analysis. In the current macroeconomic literature there is a large interest in the volatility of productivity. Standard business cycle models are often driven by exogenous productivity shocks (e.g., Smets and Wouters 2007). Further, a new literature on uncertainty shocks is pointing to the ex ante uncertainty that firms face about future operating conditions when making investment decisions (e.g., Bloom 2009). In some empirical applications, sometimes the volatility is calibrated using evidence from cross-section dispersion, which to our view is not appropriate. Of course, optimal forecasts of future volatility may contain information derived from a cross section of historical volatilities (see, e.g., Senga 2015). For the remainder of this chapter, we will focus on measures of dispersion rather than on volatility. However, we will address the possibility of cyclicality of productivity dispersion and its causes and implications.

Another issue in understanding the distribution of productivity relates to how the observations derive from a data-generating process. If the data set is a census of all existing firms, then the underlying interpretation of a statistical sampling from a probability distribution does not make sense.[4] In this case, and absent pure measurement error, the estimate of dispersion of productivity across firms in an industry should not be considered a random variable, but rather an actual measure without confidence bounds.

Even with census data, the dispersion measure becomes a random variable if one makes another interpretation of the probability space (or data-generating function). For example, firms may get a (persistent) draw from a probability distribution at entry. In this case, the observations on productivity of firms by entry-cohort could provide (p. 597) information on the underlying (time-varying) distribution from which a firm's productivity is drawn. Other possibilities include measurement error in outputs and inputs that are the underlying cause of dispersion in observed productivity. In the section on empirical dispersion measures, we will provide an overview of the types of data-generating processes that may be underlying observed productivity dispersion.

18.2.2. Economic Issues

In this section we adapt the framework of Syverson (2011) to discuss factors that affect ω_{it}, or the (relative) efficiency. Syverson distinguishes factors that operate "within" firms, or things that firms can do to change their (relative) productivity over time, and "between" factors, or things beyond a firm's control that alter a firm's relative productivity. In the following, we provide a brief overview from the recent literature to most of Syverson's factors. We exclude the factor of market competition from this list, as we see that as one of the factors that shapes the observed dispersion through allocation and selection mechanisms.

An easy way to think about, or model, heterogeneity in productivity across firms in an industry is to assume that firms receive a random draw from some underlying distribution of productivity. An interpretation of this could be that a firm has a manager or owner whose quality is random, as in Lucas (1978). The success of management may reflect differences in individual skill or the quality of practices (coordination, allocation of the labor force, etc.). Less is known about how managers actually allocate their own time, incentivize their workers, or manage relationships outside the firm. Existing papers in this context typically focused on single-industry or single-firm data, which is not surprising because these inquiries require very detailed information.[5] A nice example of this work can be found in Bloom et al. (2016) or Bushnell and Wolfram (2009). Also, the quality of management could affect the productivity of a firm over time, leading to persistence in the effect of an initial good draw. Lazear (2000) and Ichniowski and Shaw (2003) investigate management practices such as pay-for-performance schemes, work teams, cross-training, and routinized labor-management communication in forming productivity.

Rather than assuming a random draw to (a persistent component) of productivity, firms can undertake explicit actions that result in heterogeneous productivity across firms. In a simple version, firms pay a fixed (entry) fee to receive a draw from a productivity distribution (as in Hopenhayn 1992). Alternatively, firms could undertake investment in research and development (R&D), or other intangible capital. A large literature exist on the effects of IT investment on productivity (dispersion). For example, Bartelsman et al. (2017) show how use of broadband Internet is correlated with the dispersion of productivity across firms in an industry.

The literature also provides mechanisms that alter the relative position of firms in the productivity distribution, either through explicit firm decisions or through external effects such as knowledge spillovers. Bartelsman et al. (2008) analyze push-and-pull (p. 598) effects, where productivity spillovers from frontier knowledge can contribute to changes in relative productivity. Some key papers are Moretti (2004), who looks at the role of skilled workers to benefit from spillovers, and Bloom et al. (2013), who look at positive knowledge spillovers as well as business stealing effects. This last idea ties in with our next section, where we look at how interactions between agents in markets may affect the observed distribution of productivity.

The market environment for inputs and outputs conditions the decisions made by producers that can influence their productivity, as described in the preceding. The market environment also shapes the allocation of inputs across firms and the share of production and sales of each firm in the market. Competition will drive market shares toward more efficient producers, shrinking relatively high-cost firms/plants and opening up room for more efficient producers. Intra-market competition has been studied in many papers. Syverson (2004) looks at the ready-mix concrete industry (homogenous product, substitutability, etc.). International trade is another area where competition can be productivity enhancing, partly through changes in dispersion (see, e.g., Eaton and Kortum 2002; Melitz 2003; Wagner 2007). In many firm-level trade models, opening up to trade increases the “threshold” productivity below which firms exit the market, thereby reducing dispersion.

18.3. Productivity Measurement

Productivity is simply a measure of output per unit of input. With a single homogeneous output and a single homogeneous input, productivity is a cardinal number with dimensionality units of output per unit of input. With multiple inputs or output, or when inputs or outputs are not strictly homogeneous across firms or over time, typical index number issues arise. The approach then is to either define an index that meets certain desirable properties (axiomatic approach) or that can be derived from a theoretical model (see Diewert and Nakamura 2003). A productivity index then is defined as productivity relative to some reference level, for example relative to a base period of the same production unit, or relative to some other production unit. In the frontier approach, productivity of a firm is measured relative to the frontier of production possibilities (see, e.g., Chapters 2 and 4 in this *Handbook*). Essentially, productivity is a distance measure. The distribution of productivity across firms or over time should be interpreted as showing the distribution across all production units of the distance in terms of productivity between that observation and some fixed reference observation. More generally, in the empirical literature on plant-level productivity, it is customary to sweep out industry and time effects, so the productivity observations show the distance relative to the industry- and time-specific average.

Using a rather generic notation (see, e.g., Fried et al. 2008), which we will detail later as needed, production takes place by transforming a vector of inputs $\mathbf{x} \in \mathbb{R}_k^+$

into a vector (p. 599) of outputs $\mathbf{y} \in \mathbb{R}_\mathbf{m}^+$, with + and *m* aligned. This transformation takes place through a production function that defines transformation as $T(\mathbf{x}, \mathbf{y}) = 0$. Using this style of notation, one can define the inputs requirement set $S(\mathbf{y})$ with all feasible input vectors $\mathbf{x}$ that can achieve a certain output $\mathbf{y}$ (with free disposal). One can also define an isoquant

$$I(\mathbf{y}) = \{\mathbf{x} : \mathbf{x} \in S(\mathbf{y}) \text{and } \theta\,\mathbf{x} \notin S(\mathbf{y}) \quad \text{if } 0 \le \theta < 1\}$$

(or more stringently an efficient subset in case the isoquant is not strictly convex) showing the boundary of the input requirements for the given output. If one can scale down the inputs usage along a ray to the origin (in the positive orthant in input space), then the input is not technically efficient. The scalar ($\theta < 1$) needed to scale the input to the technically efficient frontier is called the *measure of technical inefficiency*. The measure of technical efficiency is then given by $\Omega(\mathbf{y}, \mathbf{x}) = \min\{\theta : \theta\,\mathbf{x} \in \mathbf{S}(\mathbf{y})\}$. A geometrically similar discussion can be made to give the distance between the output actually produced at input $\mathbf{x}$ and the technically efficient output given by the isoquant on which $\mathbf{x}$ lies.

Figure 18.1 illustrates the productivity and efficiency concepts. Starting with the narrative of frontier firms and inefficient firms, an inefficient firm using aggregate inputs $F(\mathbf{x}) = F(x_1, x_2)$ could produce higher output given its input quantities. The arrow on the left panel represents the input efficiency measure, which says given output $\mathbf{y}$, what fraction of the inputs would be needed if the firm were operating efficiently. The Farrell input efficiency measure is the ratio of the norm of the ray input vectors $\theta = |\mathbf{x}'|/|\mathbf{x}|$.[6] The horizontal arrow on the right panel shows the reduction in the aggregate input index in order to achieve efficient production.[7] Assuming scalar output, the vertical arrow on the right panel shows that the output inefficiency measure of this firm is $y'/y = \Omega'/\Omega = \theta_y$, namely, given the input vector $\mathbf{x}$, how much less is produced than the frontier firm could have produced with these inputs.

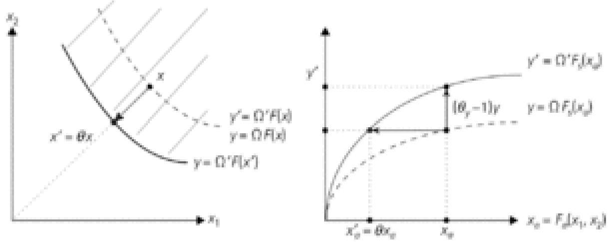

Figure 18.1. Productivity concepts. Left panel: input inefficiency; right panel: input and output efficiency.

In empirical studies using business survey data, it generally is the case that output is a scalar, y. Assume production (with free disposal) takes place according to $y \leq F(\mathbf{x}; \beta)$, (p. 600) where F is an appropriately redefined input aggregator, derived from the transformation function T noted earlier, with estimable parameters β. This can be rewritten as $y = F(\mathbf{x};\ \beta)\Omega$, where $\Omega \leq 1$ is a Farrell-type measure of inefficiency. Productivity is thus essentially the ratio of output to aggregated inputs,

$$\Omega = y/F(\mathbf{x};\ \ \beta)$$

Measurement of productivity depends on measurement of output and inputs. It also depends on specification and parameterization of the production (or input aggregator) function F and finally depends on the assumptions about the nature of the error or residual term in estimation or computation (see the review of Hulten 2001). We will address those issues most relevant to generating productivity dispersion measures from firm-level or plant-level longitudinal data.

18.3.1. Measurement of Outputs and Inputs

18.3.1.1. From Observed Data to Outputs and Inputs

A number of measurement issues need to be considered when one wants to construct productivity measures from observable data. Survey data typically record annual flows of expenses or income in currency units. The standard empirical approach is to deflate revenues or intermediate input purchases using industry-level deflators, owing to lack of product-level or firm-level prices. One consequence of this procedure is that in the presence of product differentiation the effect of heterogenous product prices is ignored. As will be discussed later in an overview of “revenue” and “quantity” total factor productivity (TFPR and TFPQ) measures, recent research has started to analyze the effects of (lack of) firm-level prices on productivity and dispersion.

Nominal output measurement generally starts with nominal sales, as recorded in firm-level survey (or register-based) data. When considering production as a physical transformation of material inputs, using capital and labor, often goods purchased for resale are subtracted from nominal sales to get a measure of output or production.[8] Sometimes data on resales are not available, but the measures for nominal output, value added, and intermediate purchases should be consistent, so that goods purchased for retail are either included in both output and intermediates, or are excluded from both. Nominal value added is then measured as output minus intermediates.

Labor input is usually measured as the number of employees or full-time equivalent (FTE) employees. If feasible, allowance should be made for work done by proprietors or unpaid family members. Often this can be proxied by adding one worker to every firm in the data set. Heterogeneity in worker quality can affect productivity measures. Sometimes wage expenditures are used as a proxy for quality-adjusted FTE, reflecting the view that this variable captures changes in the skill composition or the quality of the plant's labor force. Recent empirical work with linked employer–employee data, together with assumptions on matching/sorting between workers and firms, has made (p. 601) progress in parsing out firm-level productivity from worker heterogeneity (see, e.g., Lentz and Mortensen 2005).

Proper productivity measurement requires quality-adjusted capital service flows. This is quite difficult to measure. Survey and census data usually contain information only on the book value of the capital stock. Following procedures outlined by OECD (2009), researchers often use book values deflated by an industry-level investment deflator to proxy for capital. If firm- or plant-level investment is observable, then researchers apply some variant of the perpetual inventory method (PIM). PIM is a recursive procedure in which a deflated value of current investment is cumulated on the depreciated capital stock. Both approaches have drawbacks. First, deflated book values might be poor approximations of replacement values. Next, accumulated deflated investment may deviate from quality-adjusted service flows, which is the appropriate concept for capital input in production functions. One reason for this is a lack of proper deflators and/or lack of composition of investment that results in the proxy for capital input to be heterogeneous across plants. Further, the PIM requires an estimate of the initial capital stock, as well as estimates of depreciation by asset type at the firm level, which are not observed.

18.3.1.2. Omitted-Price Bias: Physical Productivity (TFPQ) and Revenue Productivity (TFPR)

Firm-level datasets rarely contain information on plant-level output prices and/or quantities. To obtain a plant-level output measure from nominal sales, a typical method in the empirical literature is to deflate sales using industry-level deflators. The resulting productivity index is a revenue-based indicator. Only under the assumption that the output of the industry is homogenous does TFPR calculated in this manner correctly measure productivity in quantities or physical productivity (TFPQ), and therefore technological differences across firms. If this assumption fails because products are differentiated or firms exercise market power, additional biases may result because the error term includes the effect of product prices.[9] Analogous arguments can be made about the effects of unobserved input prices because most firm-level surveys record only the total cost of inputs and not their quantities. These issues are well understood in the literature; the interpretation of alternative revenue productivity measures that emerge from various estimation procedures have become important, especially in light of the insights in Hsieh and

Klenow (2009). Keeping this in mind, we will focus on the consequences that omitted output prices may have on measuring productivity dispersion.

The difference between TFPR and TFPQ impacts the interpretation of results. Klette and Griliches (1996) show that if firms operate in an imperfectly competitive environment with heterogenous product prices, iso-elastic demand, and Cobb-Douglas technology, scale estimates from these regressions of deflated sales should be considered as a mixture of the true scale elasticity and demand parameters. The basic insight is the following. If a firm experiences a negative cost shock, or equivalently a positive productivity shock, it can increase its market share by undercutting its competitor's price. Such (p. 602) negative correlation between productivity and prices is a result of downward sloping demand. Since the increase in output is larger than the increase in sales, replacing output with revenue as the dependent variable in a least-squares regression implies that the coefficients are downward-biased estimates of the true elasticities. This also implies the dispersion of revenue productivity is smaller than that of physical productivity or efficiency. Foster et al. (2008) and Foster et al. (2017) offer empirical evidence supporting this finding. They also highlight that demand shocks exhibit high dispersion relative to physical productivity dispersion. As such, dispersion in TFPR likely reflects both dispersion in TFPQ and in demand shocks.[10]

As mentioned earlier, not accounting for product (and price) heterogeneity within industries affects productivity estimates. In the absence of data on plant-level prices and/or quantities, earlier research used the following approach addressing this issue. Assuming some structure about demand, firm-specific product prices can be substituted out from the revenue equation. Studies differ along these assumptions and have used different variables to control for firm-level prices, but it is common to assume that the firm's residual demand is iso-elastic, and that it is determined by aggregate demand and the firms' market share, which in turn is determined by the substitution effect across products within the industry. This demand structure, together with Cobb-Douglas technology, though admittedly restrictive, has the analytical advantage that it implies a closed-form solution for TFPR, regardless of assumptions about returns to scale (see Foster et al. 2016b for details). In addition, the revenue function will include a measure of industry-level output, or aggregate demand, implying that the joint estimation of demand parameters and revenue function coefficients allows the identification of factor elasticities and returns to scale. Obtaining factor elasticities in this framework is straightforward: one has to rescale the revenue elasticities and TFPR using the markup, where the markup is estimated jointly with revenue elasticities and TFPR. If our data contain information on prices and/or quantities, then combining such a demand system with a production function also allows us to identify TFPQ shocks at the plant level.[11]

18.3.1.3. Output: Gross Output or Value Added

Firms produce output using the primary inputs of capital and labor, as well as purchased materials and services. Nonetheless, in more macro-based literature, productivity often is measured on the basis of value added, starting with the work of Cobb and Douglas (1928). This approach can be motivated by the fact that in the overall economy, aggregate final demand equals aggregate value added, giving a value-added based productivity measure an intuitive interpretation. Basu and Fernald (2002) show conditions under which changes in a slightly modified aggregate value-added-based Solow residual actually measure changes in welfare, even when measured productivity and technology differ owing to various market distortions. Under perfect competition and constant returns to scale, the rate of change of value-added-based productivity is a valid measure of technical progress. Bruno (1978) provides conditions for the existence of a value-added production function, and conditions when value-added-based marginal products correctly measure true marginal products.[12]

(p. 603) In aggregating from the firm level to the aggregate level, Basu et al. (2009), using results from Basu and Fernald (2002), show a decomposition of growth of the aggregate Solow residual into terms related to growth in aggregate primary inputs, reallocation terms, and aggregate change in technology. In this aggregation, a switch is made from viewing aggregate value added as a sum of growth in final demands by product, into a sum of growth in income earned on primary factors (value added here equals gross output minus intermediate purchases) across producers. In aggregation, and abstracting from price changes, the divisia weights differ between the two, namely shares of final demand and shares in primary factor income, respectively. The latter does not have an obvious theoretical foundation.

Instead, in a productivity aggregation framework developed by Domar (1961) and expanded by Hulten (1978), an economy is viewed as a collection of firms that make products (commodities) using primary inputs and purchased commodities, and sell the products to other producers and to final demand. Within this framework, the definition of productivity and the manner in which to aggregate now depends on the level from which one is aggregating and the level to which one is aggregating. For a production unit at any level of aggregation, productivity growth is defined as growth of *net output*, or product sold to agents outside the production unit, minus (cost-share weighted) growth in primary inputs and products purchased from agents outside the production unit. In aggregation, the *Domar-weight* is the share of net output of the production unit divided by the net output of the unit to which one is aggregating. The sum of these weights is larger than one.

For example, in aggregating firm-level productivity to productivity of the (closed economy) as a whole, one defines output of a firm as total production minus own-product used in production as net output and nonprimary inputs as inputs purchased from outside the firm. For most firms, these equal gross output and intermediate input, respectively, but at farms or energy-mining firms a significant share of firm production is “produced and consumed” and needs to be netted out. For the firm level, a net-output productivity measure thus is appropriate, but in practice will equal the gross output productivity measure. In aggregation, the net output of the “whole economy” equals the sum of all firms’ net output minus the sum of all purchased inputs, in other words aggregate final demand or value added. In this case, the aggregation weights to sum each firms’ productivity growth to compute aggregate productivity growth is given by firm-level gross output divided by aggregate value added.

In an equivalent manner, one can aggregate firm productivity to the industry level, or industry-level productivity to total private nonfinancial sector, by appropriately defining net output productivity for the disaggregated units, and using Domar-weights computed by dividing net output of a disaggregate unit to net output of the unit to which one is aggregating. Corrado et al. (2007) provide a convenient notation that displays the generic properties of such aggregation, as well as expanding the concepts to include imports and exports.

Turning to productivity dispersion measures, one can in principle compute the dispersion of value-added-based productivity across firms in an industry. Nonetheless, (p. 604) even if the separability conditions needed for the value-added production function hold, gross output may be preferable because there is a market with supply and demand for output, while no market for value added exists. And, it is precisely for understanding the dynamics of such markets that productivity dispersion is interesting. With representative firms, dispersion in either value added or gross output productivity would not exist. Empirically, choosing value added instead of gross output as the dependent variable has a large effect on within-industry dispersion measures. Tables 18.2 and 18.3 offer evidence that value-added-based dispersion is much larger than output-based dispersion. This is not surprising if one considers that, to a first approximation, value-added productivity is equal to gross output productivity times the reciprocal of the share of value added in gross output.

18.3.2. Estimating the Input Aggregator

As described earlier, we can compute TFP as the ratio of output to weighted inputs, with weights as estimated in the empirical production function literature.

$$Y_{it} = K_{it}^{\beta_K} L_{it}^{\beta_L} E_{it}^{\beta_E} M_{it}^{\beta_M} \Omega_{it}$$

(18.1)

where Y, K, L, E and M denote output, capital stock, labor, energy, and material inputs, respectively. i and t index plants and time periods, and the βs denote the elasticity of Q with respect to factor inputs. It is then straightforward to define TFP as a ratio of output and an index of inputs $TFP \equiv \Omega_{it} = Y_{it}/(K_{it}^{\widehat{\beta}K} L_{it}^{\widehat{\beta}L} E_{it}^{\widehat{\beta}E} M_{it}^{\widehat{\beta}M})$. The input index is a weighted average of primary-input factors where the $\widehat{\beta}s$ are the estimated elasticities of output with respect to the appropriate input factor. A few issues are relevant in estimation of the production function or input aggregator. We start with the issue of endogeneity of a firm's factor input decisions in response to firm productivity, and discuss semi-parametric and parametric estimation methods. We also consider growth accounting methods to aggregate inputs and generate residuals, and finally refer to nonparametric data envelopment analysis (DEA)–type methods for computing productivity.

18.3.2.1. Endogeneity of Input Decisions

The following section briefly revisits estimation issues. Some of them have been analyzed in great detail in the literature, others were investigated more recently. Since productivity estimation requires elasticities in order to be able to calculate the weighted input index and compute productivity, we will use the terms *production function* and *productivity* interchangeably.

Perhaps the most extensively analyzed econometric issue is the endogeneity of production factors and unobserved TFP. As first pointed out by Marschak and Andrews (1944), least-squares-based production function estimates are rendered biased because plants consider their productivity in input decisions, but plant-level TFP is (p. 605) unobserved to the econometrician and therefore TFP is incorporated in the error term. Parametric and semi-parametric methods were developed in order to control for the variation in unobserved TFP. Parametric approaches such as instrumental variables techniques or stochastic frontiers do not explicitly control for the effects of unobserved TFP. Instead they rely on assumptions about the time series properties of plant-level productivity and apply data transformations to remove its effect from the estimating equation.[13] The aforementioned methods are all projection-based in the sense that regression techniques are used to estimate elasticities and calculate the productivity residual. Other methods, (cost-share-based techniques or growth accounting [GA] after the seminal work of Solow 1956) calculate productivity directly from data relying on first-order conditions derived from either profit maximization or cost minimization. Since GA is a nonstochastic method and therefore projection-based procedures cannot be used, the aforementioned endogeneity issue is irrelevant. However, other types of specification error do emerge if the first-order conditions are violated, for example if firms face frictions in adjusting inputs. Nevertheless, their popularity provides justification to include a short description for completeness.

18.3.2.2. Semi-Parametric Estimation

The original idea of using firm-level proxies in production function estimation was developed in Olley and Pakes (1996) (OP hereafter) in order to analyze the dynamics of the US telecommunications equipment industry. OP take account of the previously mentioned endogeneity problem by including an investment proxy in the estimation process. Assuming that investment is a monotonic and increasing function of productivity and that productivity is the only unobserved state variable, including investment in the estimation as a proxy for unobserved TFP developments allows the variation in investment to be used to infer plant-level TFP shocks. The algorithm consists of two steps. The first step provides consistent OLS estimates of variable input elasticities because the proxy controls for plant-level TFP shocks during the estimation procedure. The coefficient of capital is identified in the second step by forming moment conditions using the innovation component of TFP and lagged capital values.[14]

OP use the firm-level time series of investment to proxy for unobserved productivity. There is ample evidence that plant-level investment is lumpy. Lumpiness means that bursts of investment activity are followed by inactive periods where observed net investment is zero. It is a consequence of the presence of non-convexities in capital adjustment. Unfortunately, observations with zero investment are not informative for inferring productivity and are dropped, which may negatively affect precision if truncation significantly decreases sample size. In addition, OP works only if we observe both entrants and exiters.[15] In order to eliminate the efficiency loss caused by dropping zero-investment observations, Levinsohn and Petrin (2003) (LP hereafter) advocate the use of intermediate input cost or electricity instead of investment. LP discuss the conditions that must hold if the intermediate input is to be used as a proxy. The basis of the argument is that if intermediate inputs are less costly to adjust than investment, they are (p. 606) likely to respond more to productivity shocks. This is especially relevant in the presence of non-convexities in capital adjustment. LP also highlight that firms almost always report positive use of these variables in their data, implying that truncation due to zero proxy values is less severe.

The identifying assumptions regarding the timing of plants' input decisions have been criticized by Ackerberg et al. (2015) (ACF). ACF highlight that the optimal labor allocation is also a deterministic function of TFP and therefore the elasticity of labor is not identified. They propose a hybrid approach and offer structural assumptions on the timing of decisions concerning firms' input choices. They approach the identification problem by applying a two-step procedure that estimates all the elasticities in the second stage. Wooldridge (2009) proposed to circumvent the identification problem by estimating all the coefficients in a single generalized method of moments (GMM) step and using earlier outcomes of both capital and variable inputs as instrumental variables. His approach is advantageous because it is robust to the ACF critique and because the efficiency loss due to two-step estimation is eliminated.

18.3.2.3. Parametric Estimation (IV, GMM)

Although instrumental variable (IV) techniques are used within semi-parametric approaches, we mention IV-based methods separately because these estimate the parameters of the production function without the help of specific assumptions about firms' input decisions. At the heart of IV techniques is a general error components model developed by Blundell and Bond (2000). TFP is decomposed into a firm-fixed effect and autoregressive term, which allows for firm-specific dynamics in productivity. Blundell and Bond (2000) address the endogeneity issue by differencing the estimating equation. Under the error components assumption, differencing removes the firm-fixed effect and also controls for the dynamic effects of the autoregressive component. Obtaining the innovation in the output residual in this manner supports the construction of moment conditions that can be used to consistently estimate the parameters of the production function in a single step.[16] We note that while instrumental variable methods are attractive in principle, they are not commonly used given the lack of plausible and strong instruments on a wide scale basis to cover all industries over all time periods (see Griliches and Mairesse 1998 and Blundell and Bond 2000 for more details).

18.3.2.4. Cost-Share-Based Methods, or Growth Accounting

A frequently used nonstochastic computation method is growth accounting (GA). A typical version of GA exploits the first-order condition of a decision problem where the plant minimizes production costs given output and input prices. The first-order condition of this problem is used to rewrite elasticities as respective shares of input factors in the plant's total cost. Some of the advantages of this method include the possibility to allow for plant-level heterogeneity in elasticities,[17] easy implementation, and that it is flexible about the exact shape of the production technology. Further, available (p. 607) Monte-Carlo evidence suggests that it is accurate if the data are not subject to much measurement error (see Van Biesebroeck 2007 for details). As mentioned at the beginning of this section, GA is free of statistical problems related to endogeneity and the sensitivity of estimates to sample size. In another version of GA, first-order conditions are derived from profit maximization. In this case, output elasticities are obtained as the revenue share of input costs. Using the cost share of total costs rather than of total value has the advantage that we do not require the assumption of perfect competition. This implies that another advantage of the GA-based factor elasticities using cost shares of total costs is that they are robust to alternative demand structures. As we will discuss later, this consideration becomes important if output prices are not observed in the data.

One might argue that the first-order conditions underlying this method are unlikely to hold at all points in time at plant-level. This means the elasticity estimates and the implied productivity numbers may be biased if the first-order conditions are violated. A case in point is when input markets are subject to frictions that prevent plants from adjusting labor and capital instantaneously, especially in the presence of non-convex costs. In such cases, the validity of first-order condition becomes critical. These issues are relevant for measurement purposes because the available empirical evidence suggests that the adjustment of input factors at the plant level is subject to frictions (see Bloom 2009; Cooper and Haltiwanger 2006), implying that first-order conditions are unlikely to hold for every plant in every industry and time period. It is more reasonable to expect that they hold on average across establishments and/or over time. Therefore, it is common to impose constant elasticities across plants in the same industry and/or over time. We also note that most of the alternative estimation methods assume common factor elasticities over time within the same industry.

18.3.2.5. Hybrid Approaches

Other papers combine elements of growth accounting with other approaches, which usually involves using a first-order condition together with regression techniques. Martin (2008) is a recent example where a first-order condition is combined with the control function approach. The basic insight is that under profit maximization and imperfect competition, the output price is given by a constant markup over marginal cost. As a consequence, elasticities of fully flexible inputs are obtained as a scalar multiple of the revenue share of input costs, where the multiplier is proportional to the revenue markup. For quasi-fixed inputs like capital, we should not expect the first-order condition to hold whenever a shock hits the firm because capital adjustment is subject to non-convexities. Martin

(2008) proposes to subtract from revenues what he calls "an index of variable input usage." Then, eliminating prices from the modified revenue equation using the assumed demand structure, the revenue-elasticity of the fixed input can be written as a function of the scale elasticity and the elasticity of demand, similarly to the preceding discussion. To obtain the true coefficient of the fixed input, a control function approach is applied, but including lagged firm-level net revenues to control for (p. 608) unobserved TFP. Martin (2008) provides conditions under which variable revenues are monotone in TFPR.

18.3.3. Relative Productivity

In the previous description, productivity ($\widehat{\Omega}_{it}$) is a ratio of output divided by an estimated or computed input aggregator. If we denote our index of relative (log) productivity by $\omega_{it} \equiv \widehat{\omega}_{it} - \widehat{\omega}_0$, where $\widehat{\omega}_{it}$ is the log of productivity and $\widehat{\omega}_0$ is the reference measure, then the difference between two observations $\omega_{it} - \omega_{js}$ is consistent with the distance view: $\omega_{it} - \omega_{js} = \widehat{\omega}_{it} - \widehat{\omega}_{js}$.[18] The choice of reference productivity $\widehat{\omega}_0$ thus is not relevant per se for the dispersion measure. However, in practice, the estimated or computed residual $\widehat{\omega}_{it}$ will vary across methods with different reference productivity when the estimated or computed aggregator $F(\mathbf{x};\widehat{\beta})$ differs. Foster et al. (2017) demonstrate that empirical differences across the estimated input aggregators often imply numerical differences in both $\widehat{\omega}_{it}$ and its dispersion. Foster et al. (2017) report average dispersion from productivity distributions for 50 industries where dispersion varies between 0.24 and 0.40.[19] This range reflects nontrivial differences across estimation methods. However, all methods yield results suggesting large productivity differences across establishments. Thus for individual industries, cross-method variation in dispersion methods may be larger.

18.3.3.1. Stochastic Frontiers

In stochastic frontier production functions, the residuals are relative to a frontier firm and thus fit into the narrative of "X-inefficiency" of Leibenstein (1966). This approach, in the spirit of the production frontiers in Farrell (1957), was first developed by Aigner et al. (1977) or Meeusen and van Den Broeck (1977). The main assumption underlying this approach is that residuals can be decomposed into two components with known distributional properties. The first component, labeled "efficiency," is assumed to follow a truncated, or one-sided, distribution. The second component is "measurement error" and hence is assumed to be symmetrically distributed around the frontier. This component is considered to occur through random fluctuations outside of the firm's control. To identify the two error processes, one must make assumptions regarding independence between the two, and that both are iid across observations. More important, the two-sided error must come from a symmetric distribution with mean zero. Usually normal, $N(0,\sigma)$ and half-normal, $N^{+}(\mu,\sigma_{+})$ distributions are chosen, with error parameters estimated along with production function parameters. Loosely speaking, any skewness in errors is attributed to inefficiency, while the symmetric part can either be measurement error or across-firm (and time) heterogeneity in productivity.

For the purpose of measuring across-firm dispersion of productivity, one starts with the residuals from a production function estimation. From stochastic frontier (p. 609) estimation, the estimated residual consists of two components, $\widehat{\widehat{\omega}}_{it} = \hat{\nu}_{it} + \hat{\epsilon}_{it}$. In the frontier literature, only one component measures productivity, but we do not want to impose this interpretation, and consider ω_{it} as the building block for dispersion measures. In analysis, one can try to find a way to parse out what proportion of the dispersion can be attributed to measurement error.

In this sense, dispersion of the productivity distributions from production function versus stochastic frontier estimation will differ only owing to different parameter estimates, $\widehat{\beta}$, resulting from the different estimation procedures. While Foster et al. (2017) offer evidence that the productivity ranking across firms and dispersion results are affected by the estimation method in the context of regression-based techniques and cost-share-based procedures, there is not much evidence to date on how these results compare to frontier methods. The available evidence is presented in the next section.

18.4. Dispersion Measures

Let $\widehat{\omega}_{it} = \ln Y_{it} - \ln F(x_{i,t}\widehat{\beta})$ denote the log productivity level for establishment (firm or decision-making unit) i in time t.[20] The basic building block for our dispersion measure is the log of productivity relative to a reference measure, as described in section 18.3.3, namely $\omega_{it} \equiv \widehat{\omega}_{it} - \widehat{\omega}_0$, where $\widehat{\omega}_0$ is the reference measure.

Dispersion is related to the "width" of the productivity distribution, and generally is measured using the standard deviation (σ) or the interquartile range (iqr) measure, and thus has the same dimensionality as the underlying measure. In practice, quantile-based measures such as the interquartile or -decile range are usually preferable because they are robust to outliers. The two measures are given by

$$iqr_t(\omega_{it}) = p_{75}(\omega_{it}) - p_{25}(\omega_{it})$$

(18.2)

and

$$\sigma_t(\omega_{it}) = \left(\frac{1}{N_t}\sum_i (\omega_{it} - \overline{\omega}_t)^2\right)^{1/2}.$$

(18.3)

A time series of dispersion, either standard deviation, σ_t or interquartile range, $\mathbf{iqr}_t$, can be computed for any grouping of firms for which comparing productivity levels makes sense. In practice, estimation of input aggregators and firm-level productivity is done at the most detailed level of industry disaggregation for which enough firms are available.[21]

As mentioned in section 18.2.1, other second moments of the empirical distribution on ω_{it} can be considered; for example, for each firm we can compute the standard deviation of the productivity measures over time, $\sigma_i(\omega_{it}) = \left(\frac{1}{T_i}\Sigma_t(\omega_{it} - \overline{\omega}_i)^2\right)^{1/2}$, and we can (p. 610) call this volatility. Industry volatility could then be computed as a (size-weighted) average of firm-level volatility of firms in the industry.

Consider the following two-step procedure. First calculate iqr_{jt} for each industry j and time period t. Next, compute the average across the $j = 1 \ \ldots \ J$ industries for each t as $\overline{iqr}_t = \frac{1}{J}\Sigma_j iqr_{jt}$ and over time as $\overline{\overline{iqr}} = \frac{1}{T}\Sigma_t \overline{iqr}_t$. In this approach, each industry- and time-specific iqr_{jt} is assigned equal weight. Since industries often differ in terms of the number of plant-year observations, it is reasonable to apply a weighting scheme that accounts for such differences. For example, Foster et al. (2017) report weighted average dispersion measures in manufacturing industries where the weights are based on the number of plant-year observations in industries. This approach amounts to pooling normalized establishment-level productivity measures ω_{it} from all j and t and calculating dispersion in a single step. In our notation, their approach can be illustrated by re-indexing establishments and industries. In each t, the index of establishments is defined as $s_t = (1 \ \ldots \ N_{1t}, 1 \ \ldots \ N_{2t}, \ \ldots, \ 1 \ \ldots \ N_{jt}, \ \ldots, \ 1 \ \ldots \ N_{Jt})$, and the following vector shows indices used in the weighted average formula: $s = (s_1, s_2, \ \ldots, \ s_T)$. Assuming the panel of industries is balanced, the pooled distribution has $\Sigma_{t=1}^{T}\Sigma_{j-1}^{J}N_{jt}$ observations in total, and more populous industries will be represented according to their frequency weight $\frac{\Sigma_{t=1}^{T}N_{jt}}{\Sigma_{t=1}^{T}\Sigma_{j-1}^{J}N_{jt}}$. In this notation, $iqr(\omega_s) = p_{75}(\omega_s) - p_{25}(\omega_s)$ is equivalent to calculating the frequency-weighted average of the industry- and time-specific dispersion measures iqr_{jt}. This approach reflects the view that the different realizations of plant-specific productivity processes are outcomes of the same data-generating process and does not distinguish between the concepts of time-series volatility and cross-section dispersion. Such a procedure is appropriate in situations when time-series volatility is dwarfed by differences across plants, as is overwhelmingly the case in empirical micro datasets.

18.4.1. Empirical Evidence

The majority of previous studies focused on within-country differences across industries or sectors (see Syverson 2011 for a survey of the literature from the past decade). The main conclusion from these studies is that productivity differences across establishments are large, even within narrowly defined industries. This chapter adds another important dimension to the evidence: we compare measures also across European countries and the United States. Results on US industries are taken from Foster et al. (2017), while European dispersion statistics are based on our own calculations using data from Bartelsman et al. (2018a) and Lopez-Garcia and di Mauro (2015).

Although both the EU and US results are based on individual producers, a few qualifications are in order when comparing the results. First, US data on inputs and outputs are defined at the establishment level, while the unit of observation is a "firm" in European countries, namely the smallest production unit with independent accounting data. Second, the European data are available both for samples of firms where the number (p. 611) of employees is greater than 20 and for samples including all firms, while the results in Foster et al. (2017) are computed excluding the smallest, single-unit establishments. Third, Foster et al. (2017) use the 50 most populous manufacturing industries for estimation reasons. In contrast, the two European data sets are comprehensive in industry composition. We want to highlight that the European data allows us to estimate dispersion also in services, which is an important contribution because typical empirical data sets used in the literature contain information only on manufacturing firms.

Fourth, it is worth mentioning that although dispersion results are reported at the country level, the underlying dispersion measures are generated at a different industry detail. US results were drawn from four-digit industries, while European data, especially for small countries, allows calculations only within two-digit industries. This difference highlights an important trade-off that most researchers encounter in empirical productivity research. On the one hand, it is essential to assume some degree of homogeneity in production function coefficients in order to be able to estimate them using statistical methods. This consideration implies that it may be useful to pool industries if the narrowest industries do not have sufficient number of observations. On the other hand, possible differences in establishment-level production technology that are uncontrolled for in the estimation process affect coefficients and therefore dispersion results. For the purposes of this chapter, we assume that all characteristics relevant for productivity estimation are subsumed in production function parameters.

Fifth, we mention that differences in statistical practices across countries likely influence comparisons of dispersion measures. In recent research, Foster et al. (2017) find that dispersion measures are larger when controlling for the degree of imputed data in the sample used for computation. White et al. (forthcoming) use classification and regression trees (CART)-based methods, to estimate the empirical effects of imputation and find that underlying dispersion may be higher. No exploration has yet taken place on the differences in imputation methods across the samples in the European Union, and their effects on measured dispersion.[22]

Finally, a related issue is how dispersion measures can be made less sensitive to measurement error. In practice, outliers affect measured dispersion via two channels. First, they may affect elasticity estimates if the homogeneity assumptions about elasticities are not consistent with the data, that is, if the observations used to estimate elasticities are not derived from the same production technology. Second, extreme $\widehat{\omega}_{it}$-observations directly generate large dispersion measures. In order to reduce sensitivity to outliers, observations in the United States are filtered by output-to-capital and output-to-labor ratios using the so-called Chebyshev method. In the European Union, outliers are first filtered by trimming the 1% tails of residuals from Cobb-Douglas production functions, with final productivity estimation done on the trimmed sample.

With these considerations in mind, we now turn to empirical results. The main finding is that estimated dispersion is qualitatively similar across countries, sectors, and estimation methods, namely the difference in measured productivity within-industry is always large. For example, the first two columns in Tables 18.1 and 18.2 show that the interquartile range (IQR) of value-added based TFP varies between about 0.5 and 1.0 in (p. 612) European manufacturing industries. The comparable estimate from the United States is approximately 0.7; see the first entry in Table 18.3. IQR measures in services fall in the range between 0.52 and 1.23 in European countries (column 5 in Tables 18.1 and 18.2)—suggesting that there is nontrivial heterogeneity across sectors. Unfortunately, we do not have results from US service industries.

Table 18.1 Dispersion in (Log) TFP, European Union (ECB) (2002–2012)

	Manufacturing				Services			
	IQR		SD		IQR		SD	
	ALL	20+	ALL	20+	ALL	20+	ALL	20+
Belgium	0.72	0.52	0.55	0.41	0.72	0.51	0.56	0.40
Estonia	0.93	0.65	0.64	0.50	1.09	0.87	0.70	0.60
Finland	0.66	0.52	0.51	0.46	0.65	0.40	0.51	0.35
France	0.49	0.48	0.39	0.38	0.52	0.48	0.43	0.38
Germany	0.68	0.66	0.48	0.51	0.73	0.64	0.56	0.52
Italy	0.64	0.55	0.48	0.45	0.70	0.56	0.52	0.44
Latvia	0.98	0.83	0.61	0.60	1.23	0.78	0.80	0.59
Poland		0.82		0.63		0.87		0.73
Portugal	0.67	0.61	0.53	0.50	0.86	0.65	0.62	0.51
Slovakia		0.75		0.62		1.02		0.74

Slovenia	0.80	0.58	0.56	0.46	0.87	0.79	0.65	0.58
Spain	0.72	0.65	0.53	0.54	0.77	0.59	0.57	0.49

Log TFP (VA-based) calculated using LP (Wooldridge). Full sample of firms (ALL) or sample of firms with 20 or more employees (20+).

Source: Calculated from CompNet Descriptives File; see Lopez-Garcia et al. (2015).

Results vary also across estimation methods; see the differences in dispersion statistics across Tables 18.1 and 18.2 and the differences across the rows of Table 18.3. The entries in Table 18.1 are based on productivity measures, which are estimated using the method proposed by Wooldridge (2009). The closest candidate for comparison with US results is LP(VA) in Table 18.3, which denotes the procedure proposed in Levinsohn and Petrin (2003) with value added as the dependent variable. Although the econometric procedure used in LP(VA) and Wooldridge (2009) are not identical, LP(VA)-based results in the United States are comparable to column 1 in Table 18.1 in the sense that the estimating equation contains the same regressand and regressors. Comparing the entries in column 1 across Tables 18.1 and 18.2 for Finland, France, Germany, and Italy shows that the estimation method may generate nontrivial differences in dispersion measures. A similar conclusion holds for US results, as well (see Table 18.3).

Table 18.2 shows results based on productivity measures that are estimated using Solow residuals or GA. The table presents variants where the dependent variable is either value added (columns 1–2 and 5–6) or gross output (columns 4–5 and 7–8). Comparing the entries in column 1 to those in column 3 suggests that gross-output-based productivity (p. 613) measures are less dispersed than value-added-based ones in European manufacturing industries. The relationship is similar in services. US results confirm this finding: the LP(VA)-row of Table 18.3 shows that value-added-based dispersion is significantly larger than other, output-based measures. This empirical finding has been established by earlier studies in the United States (see Foster et al. 2017 for more details).

Table 18.2 Dispersion in (Log) TFP, European Union (Eurostat) (2001–2010)

	Manufacturing				Services			
	GO		VA		GO		VA	
	ALL	CO	ALL	CO	ALL	CO	ALL	CO
Austria	0.56	0.52	0.20	0.19	0.76	0.75	0.38	0.38
Denmark	0.58	0.57	0.25	0.24	0.73	0.72	0.34	0.33
Finland	0.70	0.67	0.36	0.33	0.81	0.78	0.48	0.45
France	0.55	0.53	0.25	0.25	0.62	0.61	0.39	0.37
Germany	0.47	0.47	0.19	0.19	0.51	0.51	0.22	0.22
Italy	0.86	0.83	0.40	0.35	1.04	1.00	0.52	0.46
Nether-lands	0.56	0.56	0.24	0.24	0.72	0.71	0.38	0.37
Norway	0.80	0.79	0.38	0.37	0.96	0.94	0.51	0.50
Poland	1.01	0.99	0.55	0.52	1.18	1.15	0.96	0.93
Sweden	0.70	0.70	0.40	0.39	0.84	0.83	0.60	0.59

United Kingdom	0.76	0.74	0.45	0.43	0.98	0.96	0.60	0.58

Solow Residual measures of productivity, value added based or gross output based. Full sample of firms (ALL), or continuing firms (CO).

Source: Calculated from ESSNet; see Bartelsman et al. (2018).

Table 18.3 Descriptive Statistics of TFP Distributions in US Manufacturing Industries (1972–2010)

	IQR	SD
LP(VA)	0.68	0.57
LP(*Q*, GR)	0.29	0.31
GA(Q)	0.24	0.22
WLPM(*Q*)	0.40	1.88

LP(VA): using value added as the dependent variable; LP(Q,GR): using revenues as the dependent variable and grid search procedure for numerical optimization.

Source: Foster et al. (2017).

(p. 614) An interesting cross-country comparison emerges if we contrast the entries in column 3 of Table 18.2 with GA(*Q*) in Table 18.3. GA-based dispersion in revenue-productivity in US manufacturing industries (0.24) is closest to that in Germany (0.19), Austria (0.20), The Netherands (0.24), Denmark (0.25), and France (0.25).[23] Comparing numbers across the restricted and unrestricted European samples in Table 18.1 and the United States implies that excluding smaller plants (EU) or industries (US) is likely to yield smaller productivity dispersion. This finding suggests that restricting the scope of the estimation sample generally implies smaller dispersion.

A further comparison can be made of estimates for European and US manufacturing dispersion in column 3 of Table 18.2 and Table 18.3, respectively. The estimates imply that the plant at the 75th percentile of the productivity distribution in the average European industry generates between 20% and 55% more revenue using the same amount of inputs than the plant at the 25th percentile. This range varies between 24% and 40% in US manufacturing, depending on the estimation method. As we will show, the variation in dispersion may be significant along a variety of dimensions such as industries, sectors, countries, time, and estimation methods. We therefore find it remarkable that these measures are comparable in magnitude and that they all suggest that cross-plant differences in productivity are sizable.

The differences we have seen so far in these tables are country specific. However, dispersion also varies across industries and time. Explaining such variation is beyond the scope of this chapter, but the dispersion underlying the entries in our tables warrants further analysis that should explore the properties of the dispersion distribution in more detail. The existence of comparable data across countries, industries, and time of within-industry dispersion in productivity will allow for empirical explorations into correlates related to policy, institutions, and technology. Recently, several studies attempted to exploit and explain the within-country variation in dispersion. In perhaps the most popular area of application, cross-country differences in the dispersion of revenue productivity measures are associated with the degree of misallocation (see Bartelsman et al. 2013; Foster et al. 2016b; Hsieh and Klenow 2009).

We conduct a simple analysis of variance by regressing a dispersion measure on country, industry, and time effects for both the CompNet (ECB) and the ESSNet (Eurostat) panels. Table 18.4 shows the analysis of variance results for the standard deviation of TFP and the interquartile range. The main findings are the following. Country and industry fixed effects are invariably significant; they explain close to two-thirds of the variance of the standard deviation of productivity or interquartile range of productivity (see columns 1 and 3). The explanatory power of these factors is similar in the upper and lower part of the support of the productivity distribution (columns 2–3 and 5–6). The contribution of time effects is relatively small, and some preliminary analysis of the time effects does not show a clear cyclical pattern.

Our data allow us to shed more light on the potential determinants of differences in dispersion. Instead of regressing dispersion measures only on country, industry, and time dummies, we add an indicator of interest to the regression. This approach is not meant to identify an exogenous effect of an explanatory indicator. Instead, the partial correlation estimated through the regression is a useful starting point in dissecting the high explanatory power of country- and industry-fixed effects. One indicator of interest (p. 615) is the possible differences in the phase of the business cycle, which can be measured using the output-gap. This may be relevant because earlier evidence suggests that dispersion in US manufacturing appears countercyclical (see the findings in Kehrig 2015, for example). While we did not find clear cyclical patterns in the time component of the variance decomposition, further analysis may reveal how exogenous shifts in demand may affect industry dispersion. Another, largely unexplored, area that may be relevant for dispersion is related to the differences in country- or industry-specific regulations. For example, employment protection, trade regulations, and financial conditions all have been seen to affect firm and input factor dynamics. In order to understand the link to productivity dispersion, further theoretical and empirical work seems justified in this context.

Table 18.4 Variance Decomposition of Dispersion Measures

	ESSNet Data			CompNet Data		
	SD	**P90-mean**	**mean-P10**	**SD**	**P90-mean**	**mean-P10**
Country	43.4	44.9	162.7	29.6	28.8	249.3
Industry	41.2	17.1	80.4	32.8	26.4	148.4
Time	.7	.8	2.6	1.1	.9	12.6
Num. Obs.	1,964	1,949	1,948	6,288	6,288	6,288
Total SSQ	122.1	82.0	364.6	102.7	92.2	570.0

Source: Calculated from ESSNet and CompNet Data. Data described in Bartelsman et al. (2018) and Lopez-Garcia et al. (2015).

In the following we show results from a simple exercise to estimate the correlation between an indicator of technology use and dispersion. As suggested in Bartelsman et al. (2016), the intensity with which broadband Internet is used by firms in an industry is seen to be correlated with dispersion. This implies that indicators of innovation and technology intensity could control for differences in the industry-specific technology mix. We show some evidence on this from the Eurostat data that include information on technology use by firms. The results of this analysis are shown in Table 18.5. Each row displays the coefficient and t-statistic of each indicator in a regression of dispersion on the indicator and on country, industry, and time fixed effects. Our results suggest that faster growth in European industries is associated with smaller dispersion, a finding consistent with earlier results on US data in Kehrig (2015). Various indicators of technology use are positively correlated with dispersion, suggesting that more innovative/tech-using industries are also more likely to be dispersed. This result is consistent with the mechanism in which entrepreneurial innovation entails more experimentation, thereby increasing productivity dispersion compared to sectors with less innovation. This mechanism and its empirical consequences are analyzed in Foster et al (2017b), who look at the relationship between innovative activity, entry, productivity dispersion and growth. Alternatively, sectors facing large shocks in business conditions that affect measured productivity may use information and communications–related technology to reduce adjustment frictions (see Gal 2017). (p. 616)

Table 18.5 Correlates of Productivity Dispersion

	ESSNet Data	
	Coef	***t*-stat**
Industry growth	–.04	6.4
Human capital intensity	.75	10.6
IT human capital	.65	6.1
Process innovation	.08	2.4
Product innovation	.13	4.2
Organization innovation	.12	3.0
New product turnover	.20	2.9
Broadband intensity	.11	2.8
Pct ICT intensive firms	.14	2.8
Supply chain integration	–.10	2.0

Notes: Each row presents the coefficient for the indicator from a regression of the standard deviation of TFP on country, industry, and time fixed effects and the indicator. Data from ESSNet (where number of firms underlying observation > 40). The explanatory variables are indicators from the Community Innovation Survey and the ICT Use Survey that have been linked to firm-level data. Data described in Bartelsman et al. (2018).

18.5. Conclusion and Research Agenda

This chapter provides an overview of the methods currently used to construct measures of productivity dispersion using data from large, comprehensive, samples of plants or firms. In particular, the chapter draws from work done in the European Union, funded by Eurostat through the ESSNet programs[24] and by the European Central Bank's Competitiveness Network, and from work done with the Annual Survey of Manufactures and the Census of Manufactures at the US Census Bureau. The chapter further provides a comparison of estimates of productivity dispersion for different methods and for a selection of countries. Evidence also is provided on some of the correlates of productivity dispersion. Disentangling causes of productivity dispersion remains difficult and requires modeling of causes and effects, as well as empirical strategies to identify the underlying mechanisms.

The empirically observed dispersion of productivity has been an awkward fact for models of production with representative firms or for models where resources always are allocated optimally. In section 18.2, a discussion is presented of different ways in which to understand the existence of productivity dispersion. To start, many forms of measurement error could contribute to observed dispersion. Next, decisions made by (p. 617) firms that alter their productivity are a source of dispersion, as long as some form of friction is preventing instantaneous allocation of resources to the firm with the highest productivity. Finally, forces of selection and allocation tend to reduce dispersion, but may be held back by policy distortions, or by frictions in “taste and technology” (i.e., consumer learning, informational frictions, or search-and-matching processes).

Much work remains to be done to understand productivity dispersion. For example, the early attempt of Hsieh and Klenow (2009) to use observed dispersion as an indicator of misallocation of resources has run into criticism (e.g., Bartelsman et al. 2013; Bartelsman et al. 2018b; Brown et al. 2016; Foster et al. 2016b). The link between dispersion and misallocation hinges on the assumption of constant returns to scale, but also breaks down with alternative measures of revenue productivity or with alternate interpretations of estimated distortions. Further, careful measurement of dispersion in different sets of countries, industries, or time periods clearly places question marks on a simple monotonic relationship between productivity dispersion and misallocation.

The research agenda can be broken down into different themes. More work needs to be done to improve basic measurement of the underlying inputs and outputs. Linking the business surveys on production with information on the skills and education of each employee per firm, or with surveys on capital investment by type and quality, or with information on technology use or management quality, could all improve measurement of productivity. Similarly, information on the product markets and customers could help disentangle price, quality, and markups, further improving the measure of productivity. The effect of better measurement of productivity on the magnitude of dispersion across producers in an industry remains an empirical matter.

The econometric aspects of productivity measurement attract much research, as witnessed by many of the contributions to this *Handbook*. Some improvements in estimation, for example in disentangling productivity and markups, are generally expected to reduce measured dispersion (e.g., entrants may seem to have low productivity, thereby increasing dispersion, but this effect disappears once their lower-than-average markups are properly accounted for). On the other hand, accounting for statistical issues such as the presence of item nonresponse and imputation may increase dispersion.

Theoretical and empirical explorations into the sources of firm-level productivity evolution are an equally interesting area. Much work has already taken place here, for example following two disparate strands of work as described in Comin and Mulani (2009) and Acemoglu et al. (2013). Research should be partial equilibrium in nature, in the sense that it should try to isolate the sources driving (heterogeneous) productivity at the plant or firm level from market forces that select firms and allocate resources and market shares.

Beyond identifying isolated factors that drive dispersion, more work needs to be done on the implications of using heterogenous firm models in dynamic general equilibrium frameworks. These models should simultaneously take into account firm decisions that affect (future) productivity and market outcomes relating to allocation of input and output. The parameters of such models can be informed through calibration with (p. 618) moments from firm-level data sets, or can be estimated through methods of indirect inference (see, e.g., Dridi et al. 2007; Gouriéroux et al. 2010).

As measures of productivity dispersion are becoming available for researchers, systematic empirical explorations into correlates of dispersion can be made to understand how dispersion can vary across sectors, countries, or time. A recent example in this area is Kehrig (2015), who explores the differential effect that market selection mechanisms may have on dispersion over the business cycle. Another direction is taken by Brown et al. (2016), who take a theoretical and empirical look at the role of adjustment frictions on measured dispersion. The simple example given in this chapter, relating productivity dispersion in a country-industry-time panel to fixed effects and factors that vary across country or industry, could be the basis of a line of empirical literature.

This chapter serves as a guide to aid researchers in building up comparable measures of dispersion of productivity for a large set of countries, industries, and time periods. Our hope is that the availability of such data, together with research along the lines sketched in the preceding, will increase our understanding of the effect of statistical quality and methodological choices on measures of dispersion. The areas of economics where such measures can be important are wide, ranging from dynamic macro models of business cycles and growth to structural micro models of firm behavior and market outcomes. The next iteration of the *Handbook* likely will be able to host a more mature chapter on productivity dispersion, with more questions answered than asked.

References

Acemoglu, D., P. Aghion, and F. Zilibotti. 2006. “Distance to Frontier, Selection, and Economic Growth.” *Journal of the European Economic Association* 4(1): 37–74. (p. 621)

Acemoglu, Daron, Ufuk Akcigit, Nicholas Bloom, and William Kerr. 2013. “Innovation, Reallocation and Growth.” Center for Economic Studies, US Census Bureau Working Paper 13-23.

Ackerberg, Daniel A., Kevin Caves, and Garth Frazer. 2015. “Identification Properties of Recent Production Function Estimators.” *Econometrica* 83(6): 2411–2451.

Aigner, Dennis, C. A. Knox Lovell, and Peter Schmidt. 1977. “Formulation and Estimation of Stochastic Frontier Production Function Models.” *Journal of Econometrics* 6(1): 21–37.

Asker, John, Allan Collard-Wexler, and Jan De Loecker. 2014. “Dynamic Inputs and Resource (Mis)Allocation.” *Journal of Political Economy* 122(5): 1013–1063.

Bartelsman, Eric, Sabien Dobbelaere, and Bettina Peters. 2015. “Allocation of Human Capital and Innovation at the Frontier: Firm-Level Evidence on Germany and the Netherlands.” *Industrial and Corporate Change* 24(5): 875–949.

Bartelsman, Eric J., and Mark Doms. 2000. “Understanding Productivity: Lessons from Longitudinal Microdata.” *Journal of Economic Literature* 38(3): 569–594.

Bartelsman, Eric J., Eva Hagsten, and Michael Polder. (2018a). “Micro Moments Database for Cross-Country Analysis of ICT, Innovation, and Economic Outcomes.” *Journal of Economic and Management Strategy* 27(3).

Bartelsman, Eric J., Jonathan Haskel, and Ralf Martin. 2008. “Distance to Which Frontier? Evidence on Productivity Convergence from International Firm-Level Data.” CEPR Discussion Paper Series, DP7032.

Bartelsman, Eric J., Paloma Lopez-Garcia, and Giorgio Presidente. 2018b. “Cyclical and Stuctural Variation in Resource Reallocation: Evidence for Europe. Tinbergen Institute Discussion Paper 18-070/VI.

Bartelsman, Eric J., Pieter A. Gautier, and Joris De Wind. 2016. “Employment Protection, Technology Choice, and Worker Allocation.” *International Economic Review* 57(3): 787–826.

Bartelsman, Eric, John Haltiwanger, and Stefano Scarpetta. 2013. “Cross-Country Differences in Productivity: The Role of Allocation and Selection.” *American Economic Review* 103(1): 305–334.

Bartelsman, Eric, George van Leeuwen, and Michael Polder. 2017. “CDM Using a Cross-country Micro Moments Database.” *Economics of Innovation and New Technology* 26(1–2): 168–182.

Basu, Susanto, and John G. Fernald. 2002. “Aggregate Productivity and Aggregate Technology.” *European Economic Review* 46(6): 963–991.

Basu, Susanto, Luigi Pascali, Fabio Schiantarelli, and Luis Serven. 2009. “Productivity, Welfare and Reallocation: Theory and Firm-Level Evidence.” National Bureau of Economic Research Working Paper No. 15579.

Bloom, Nicholas. 2009. “The Impact of Uncertainty Shocks.” *Econometrica* 77(3): 623–685.

Bloom, Nicholas, Raffaella Sadun, and John Van Reenen. 2016. “Management as a Technology?” Stanford University Graduate School of Business Research Paper 16–27, Stanford, CA.

Bloom, Nicholas, Mark Schankerman, and John Van Reenen. 2013. “Identifying Technology Spillovers and Product Market Rivalry.” *Econometrica* 81: 1347–1393.

Blundell, R., and S. Bond. 2000. “GMM Estimation with Persistent Panel Data: An Application to Production Functions.” *Econometric Reviews* 19(3): 321–340.

Brown, J. David, Emin Dinlersoz, and John S. Earle. 2016. “Does Higher Productivity Dispersion Imply Greater Misallocation? A Theoretical and Empirical Analysis.” Center for Economic Studies, US Census Bureau Working Paper No. 16-42.

Bruno, Michael. 1978. “Duality, Intermediate Inputs and Value-Added.” In *Production Economics: A Dual Approach to Theory and Applications*, Vol. 2, edited by Melvyn Fuss and Daniel McFadden, Chapter III, 3–16. Amsterdam: North Holland. (p. 622)

Bushnell, James B., and Catherine D. Wolfram. 2009. “The Guy at the Controls: Labor Quality and Power Plant Efficiency.” In *International Differences in the Business Practices and Productivity of Firms*, edited by Richard B. Freeman and Kathryn L. Shaw, 79–102. Chicago; London: University of Chicago Press.

Cobb, Charles W., and Paul H. Douglas. 1928. “A Theory of Production.” *American Economic Review* 18(1): 139–165.

Collard-Wexler, Allan, and Jan De Loecker. 2014. “Reallocation and Technology: Evidence from the US Steel Industry.” *American Economic Review* 105(1): 131–171.

Comin, Diego, and Sunil Mulani. 2009. “A Theory of Growth and Volatility at the Aggregate and Firm Level.” *Journal of Monetary Economics* 56(8): 1023–1042.

Cooper, Russell W., and John C. Haltiwanger. 2006. “On the Nature of Capital Adjustment Costs.” *The Review of Economic Studies* 73(3): 611–633.

Corrado, Carol, Paul Lengermann, Eric J. Bartelsman, and J. Joseph Beaulieu. 2007. “Sectoral Productivity in the United States: Recent Developments and the Role of IT.” *German Economic Review* 8(2): 188–210.

Criscuolo, Chiara, Peter Gal, and Carlo Menon. 2014. “The Dynamics of Employment Growth.” OECD Science, Technology and Industry Policy Papers 14.

De Loecker, Jan. 2011. “Recovering Markups from Production Data.” *International Journal of Industrial Organization* 29(3): 350–355.

Diewert, W. Erwin, and Alice O. Nakamura. 2003. “Index Number Concepts, Measures and Decompositions of Productivity Growth.” *Journal of Productivity Analysis*, 19(2–3): 127–159.

Domar, Evsey D. 1961. “On the Measurement of Technological Change.” *The Economic Journal* 71(284): 709–729.

Dridi, Ramdan, Alain Guay, and Eric Renault. 2007. “Indirect Inference and Calibration of Dynamic Stochastic General Equilibrium Models.” *Journal of Econometrics* 136(2): 397–430.

Eaton, Jonathan, and Samuel Kortum. 2002. “Technology, Geography, and Trade.” *Econometrica* 70(5): 1741–1779.

Farrell, M. J. 1957. “The Measurement of Productive Efficiency.” *Journal of the Royal Statistical Society, Series A (General)* 120(3): 253–290.

Foster, Lucia, Cheryl Grim, and John Haltiwanger. 2016a. “Reallocation in the Great Recession: Cleansing or Not?” *Journal of Labor Economics* 34(S1): S293–S331.

Foster, Lucia, Cheryl Grim, John Haltiwanger, and Zoltan Wolf. 2017a. “Macro and Micro Dynamics of Productivity: From Devilish Details to Insights.” National Bureau of Economic Research, Working Paper Series no. 23666.

Foster, Lucia, Cheryl Grim, John Haltiwanger, and Zoltan Wolf. 2017b. “Innovation, Productivity Dispersion, and Productivity Growth.” National Bureau of Economic Research Working Paper Series No. 24420.

Foster, Lucia, Cheryl Grim, John Haltiwanger, and Zoltan Wolf. 2016b. “Firm-Level Dispersion in Productivity: Is the Devil in the Details?” *American Economic Review* 106(5): 95–98.

Foster, Lucia, John Haltiwanger, and Chad Syverson. 2008. “Reallocation, Firm Turnover, and Efficiency: Selection on Productivity or Profitability?” *American Economic Review* 98(1): 394–425.

Fried, Harold O., C. A. Knox Lovell, and Shelton S. Schmidt. 2008. *The Measurement of Productive Efficiency and Productivity Growth*. Oxford: Oxford University Press.

Gal, Peter N. 2017. “How Does Demand Volatility Encourage ICT Use?” In *Essays on the Role of Frictions for Firms, Sectors and the Macroeconomy*, edited by Peter N. Gal, 53–78. PhD dissertation, Vrije Universiteit Amsterdam. (p. 623)

Gouriéroux, Christian, Peter C. B. Phillips, and Jun Yu. 2010. “Indirect inference for Dynamic Panel Models.” *Journal of Econometrics* 157(1): 68–77.

Griliches, Z., and J. Mairesse. 1998. “Production Functions: The Search for Identification.” In *Econometrics and Economic Theory in the Twentieth Century: The Ragnar Frisch*

Centennial Symposium, edited by Steinar Strom, 169–203. Cambridge: Cambridge University Press.

Hopenhayn, Hugo A. 1992. “Entry, Exit, and Firm Dynamics in Long Run Equilibrium.” *Econometrica* 60(5): 1127–1150.

Hsieh, Chang-Tai, and Peter J. Klenow. 2009. “Misallocation and Manufacturing TFP in China and India.” *Quarterly Journal of Economics* 124(4): 1403–1448.

Hulten, Charles R. 1978. “Growth Accounting with Intermediate Inputs.” *The Review of Economic Studies* 45(3): 511–518.

Hulten, Charles R. 2001. “Total Factor Productivity: A Short Biography.” In *New Developments in Productivity Analysis*, Vol. 63 of *NBER Conference on Research in Income and Wealth*, edited by Edwin R. Dean, Charles R. Hulten, and Michael J. Harper, 1–54. Chicago: University of Chicago Press.

Ichniowski, Casey, and Kathryn Shaw. 2003. “Beyond Incentive Pay: Insiders Estimates of the Value of Complementary Human Resource Management Practices.” *Journal of Economic Perspectives* 17(1): 155–180.

Kehrig, Matthias. 2015. “The Cyclical Nature of the Productivity Distribution.” Social Science Research Network SSRN Scholarly Paper ID 1854401, Rochester, NY.

Klette, Tor Jakob, and Zvi Griliches. 1996. “The Inconsistency of Common Scale Estimators When Output Prices Are Unobserved and Endogenous.” *Journal of Applied Econometrics* 11(4): 343–362.

Lazear, Edward P. 2000. “The Power of Incentives.” *American Economic Review* 90(2): 410–414.

Leibenstein, Harvey. 1966. “Allocative Efficiency vs. ‘X-Efficiency.’” *American Economic Review* 56(3): 392–415.

Leibenstein, Harvey. 1978. “X-Inefficiency Xists: Reply to an Xorcist.” *American Economic Review* 68(1): 203–211.

Lentz, Rasmus, and Dale T. Mortensen. 2005. “Productivity Growth and Worker Reallocation.” *International Economic Review* 46(3): 731–749.

Levinsohn, James, and Amil Petrin. 2003. “Estimating Production Functions Using Inputs to Control for Unobservables.” *The Review of Economic Studies* 70(2): 317–341.

Lopez-Garcia, Paloma, and Filippo di Mauro. 2015. “Assessing European Competitiveness: The New CompNet Micro-Based Database.” European Central Bank Working Paper Series 1764.

Lucas, Robert E. 1978. “Size Distribution of Business Firms.” *Bell Journal of Economics* 9(2): 508–523.

Marschak, Jacob, and William H. Andrews. 1944. "Random Simultaneous Equations and the Theory of Production." *Econometrica* 12(3–4): 143–205.

Marshall, Alfred. 1920. *Principles of Economics.* Revised ed. London: Macmillan.

Martin, Ralf. 2008. "Productivity Dispersion, Competition and Productivity Measurement." PhD dissertation, London School of Economics.

Meeusen, Wim, and Julien van Den Broeck. 1977. "Efficiency Estimation from Cobb-Douglas Production Functions with Composed Error." *International Economic Review* 18(2): 435–444.

Melitz, Marc J. 2003. "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity." *Econometrica* 71(6): 1695–1725.

Moretti, Enrico. 2004. "Workers' Education, Spillovers, and Productivity: Evidence from Plant-Level Production Functions." *American Economic Review* 94(3): 656–690. (p. 624)

OECD. 2009. *Measuring Capital: OECD Manual*. Paris: OECD.

Olley, G. Steven, and Ariel Pakes. 1996. "The Dynamics of Productivity in the Telecommunications Equipment Industry." *Econometrica* 64(6): 1263–1297.

Robbins, Lionel. 1928. "The Representative Firm." *The Economic Journal* 38(151): 387–404.

Salter, W. E. G. 1960. *Productivity and Technical Change.* Cambridge: Cambridge University Press.

Senga, Tatsuro. 2015. "A New Look at Uncertainty Shocks: Imperfect Information and Misallocation." Queen Mary University of London, School of Economics and Finance Working Paper 763.

Smets, Frank, and Rafael Wouters. 2007. "Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach." *American Economic Review* 97(3): 586–606.

Solow, Robert M. 1956. "A Contribution to the Theory of Economic Growth." *Quarterly Journal of Economics* 70(1): 65–94.

Stigler, George J. 1976. "The Xistence of X-Efficiency." *American Economic Review* 66(1): 213–216.

Syverson, Chad. 2004. "Product Substitutability and Productivity Dispersion." *Review of Economics and Statistics* 86(2): 534–550.

Syverson, Chad. 2011. "What Determines Productivity?" *Journal of Economic Literature* 49(2): 326–65.

Van Biesebroeck, Johannes. 2007. "Robustness of Productivity Estimates." *Journal of Industrial Economics* 55(3): 529–569.

Wagner, Joachim. 2007. "Exports and Productivity: A Survey of the Evidence from Firm-level Data." *World Economy* 30(1): 60–82.

White, T. Kirk, Jerome P. Reiter, and Amil Petrin. forthcoming. "Imputation in U.S. Manufacturing Data and Its Implications for Productivity Dispersion." *The Review of Economics and Statistics*. **https://doi.org/10.1162/REST_a_00678**

Wooldridge, Jeffrey M. 2009. "On Estimating Firm-Level Production Functions Using Proxy Variables to Control for Unobservables." *Economics Letters* 104(3): 112–114.

Notes:

(1.) The authors would like to thank two referees for comments and suggestions, Bert Balk for providing guidance at an early stage of the project, and the numerous colleagues and collaborators at the European Central Bank, EuroStat, US Bureau of the Census, and many National Statistical Agencies and Central Banks in the European Union. Bartelsman would like to thank the ECB research visitor program for support under contract 24591/R/2012. This chapter has directly or indirectly made use of confidential firm-level data in accordance with all appropriate rules and regulations, and all published results meet relevant disclosure rules. The findings and opinions expressed in this chapter are the authors' alone and do not reflect policy of the US Census Bureau, the ECB, Eurostat, or any other agency involved.

(2.) Alternatively, in the empirical literature one often speaks of a "data generating process," that is, events that occur through which the outcome data are generated.

(3.) In general, one needs to take care of heteroskedasticity in measuring volatility: in practice, firm-level productivity is highly persistent, and an error variance should be estimated using, e.g., an auto-regressive process.

(4.) One could consider that all existing firms represent a draw from a distribution of firm distributions that could have existed in alternative "worlds." When analyzing within-industry distributions of firms, sample selection can be an issue even with full census data. For example, the draw of all observed firms in one country, or time period, could differ from that in another location or era because the economic environment prevents certain types of firms from entering or results in rapid death (before observation) of other types.

(5.) A recent effort to collect more detailed data from US manufacturers is through the Managerial and Organizational Practices Survey of the US Census Bureau, https://www.census.gov/mcd/mops.

(6.) We are abstracting from efficiency in factor mix. Given input prices, we can decompose the inefficiency of the original point $\mathbf{x}$ into technical and allocative inefficiency.

(7.) We have defined a scalar aggregate input index by splitting the production function such that$F(\mathbf{x}) = F_s(F_a(\mathbf{x}))$ and$x_a = F_a(\mathbf{x})$ and$F_a(.)$ exhibits constant returns to scale. Note that we have drawn$f_s()$ to have decreasing returns to aggregate input. Under constant returns, the input efficiency measure is the reciprocal of the output efficiency measure.

(8.) In our empirical section we will address the relevance of this issue, as well as the more fundamental distinction between gross output and value added measures of production.

(9.) While there are studies in which the chosen data set allows either to calculate quantity directly or to infer the effect of prices, these analyses are usually restricted to a small set of industries or even a single industry. Recent examples from this literature are Collard-Wexler and De Loecker (2014); Foster et al. (2008); Martin (2008); Syverson (2004).

(10.) In addition to differences in plants' efficiency levels and variation in product prices, other factors are potentially important contributors to TFPR dispersion. Hsieh and Klenow (2009) highlight the role of distortions in generating dispersion in TFPR. Bartelsman et al. (2013) emphasize the role of frictions such as overhead factor costs. Asker et al. (2014) explore the role of adjustment frictions in generating dispersion in TFPR.

(11.) Using data on product quotas to control for product-specific demand shocks at the plant level, De Loecker (2011) follows this thread and combines a demand system with a production function in order to recover estimates of TFPQ, since output quantities are unobserved in his data. The approach follows the line of thought of earlier papers and is extended to a case when plants produce a variety of products. In terms of empirical implementation, the main difference relative to the single-output case is that a weighted average of demand-specific aggregate-deflated revenues is included, instead of the total revenues in the industry.

(12.) Bruno (1978) explores the question of under what conditions double-deflated value added results in a production function where the partial derivatives will correctly measure the marginal productivities of production factors. Such a production function exists if the intermediate input satisfies one of three conditions: (1) it is used in fixed proportion to gross output; (2) the relative price of the intermediate inputs to value added remains constant; (3) the gross output production function is functionally separable into the intermediate and primary inputs.

(13.) The basic papers are Blundell and Bond (2000); Griliches and Mairesse (1998).

(14.) A more general point about proxy methods is related to polynomial approximations. Proxy methods use polynomials at two points of the estimation algorithm. First, a polynomial of the state variables and the proxy is included in the first step to approximate unobserved productivity. Second, to determine the expected component of TFP, its estimated value is projected on a polynomial expansion of its past values. This step is supported by a Markovian assumption about plant-level TFP. The innovation obtained in this approximation is used to construct a moment condition in order to estimate the elasticity of capital. While polynomial series provide flexible approximations, the higher order terms are also likely to exacerbate measurement error present in microdata.

(15.) OP focus on the period between the early 1970s and the mid-1980s in the telecommunications equipment industry. During this period, the industry saw large changes in the size of plants and significant entry and exit. Therefore they model plants' entry and exit decisions that depend on productivity and control for these effects in the estimation procedure. This may be an important feature of the approach in cases where the data is subject to nonrandomness. The findings in Foster et al. (2017) suggest that controlling for the effect of selection may have an effect on dispersion estimates.

(16.) Blundell and Bond (2000) provide two sets of moment conditions. The first set is based on the orthogonality of $t-3$ levels of input factors and current-period differences of output residuals. The second set is constructed using $t-2$ levels of input factors and current period levels of the output residuals.

(17.) In terms of empirical results, this assumption comes at a cost. Foster et al. (2017) offer indirect evidence suggesting that plant-level shares are likely to be noisier than industry-level shares.

(18.) This measure also is consistent with the set of index number properties proposed in Diewert and Nakamura (2003).

(19.) The range is given for dispersion of gross-output-based productivity. As discussed later, dispersion for value-added-based productivity is substantially higher.

(20.) The possible specifications for the input aggregator $F()$ are described in the previous section.

(21.) There is a trade-off here: estimation of production functions at a higher level of aggregation may introduce noise in the productivity estimates owing to imposition of common output elasticities across firms while they actually differ, but given precision of the productivity estimates, addition of more firms for computation of dispersion increases precision. In Tables 18.1 and 18.2 for EU countries, we require at least 50 firms in the industry. In calculations using data from US manufacturing, a selection is made of the 50 four-digit industries with the highest number of plant-year observations. The average number of plants (over time) in these industries varies between 400 and 3,900.

(22.) While the production data used the EU exercises based on the Structural Business Statistics surveys in each country, some countries enrich the data with information from official registers, for example on payroll tax or value added tax, and do partial imputation for missing fields.

(23.) WLP(Q) in Table 18.3 shows results obtained by the method described in Wooldridge (2009), but using output as the dependent variable. Therefore, the results are not directly comparable to the results in Table 18.1. However, earlier results in Foster et al. (2017)

suggest that recalculating Table 18.1 using output as the dependent variable would imply larger dispersion. Moreover, existing Monte-Carlo evidence in Foster et al. (2017) shows that the standard error of dispersion statistics implied by proxy methods may be large, especially when using the procedure proposed by Wooldridge (2009). This is an indication that appropriate caution is needed because these estimation methods seem to be more sensitive to sample size.

(24.) ESSLimit, ESSLait.

Eric J. Bartelsman

Eric J. Bartelsman, Vrije Universiteit Amsterdam

Zoltan Wolf

Zoltan Wolf, CES, US Census Bureau